2009

Impact of private data mining on personal privacy from agents of government

Amy Joines
Iowa State University

Impact of private data mining on personal privacy from agents of government

by

Amy Joines

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Information Assurance and Computer Engineering

Program of Study Committee:
Thomas Daniels, Major Professor
Doug Jacobson
Steffen Schmidt

Iowa State University

Ames, Iowa

2009

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank Dr. Tom Daniels for all his help while I was writing this thesis. His guidance and suggestions were instrumental in helping me complete this project.

I would also like to thank Lee and my parents, who often talked to me about the concepts and arguments included throughout this paper. Their insights and perspectives helped me reach greater depth during my analysis and helped me challenge my own assumptions and bias.

ABSTRACT

The balance of power between the individual and the government has been tipped in favor of the government. Recent advances in technology have led to the commoditization of personal information. The vast amounts of information collected and the ease with which they can be transferred to other parties has led to the rise of personal profile data mining. Due to Supreme Court decisions in the 1970s, the government can easily use or buy these profiles. Therefore, personal privacy from agents of government is being reduced due to data mining.

This paper will explore the causes, implications, and potential solutions to this problem. It will explain the value of privacy and why it should be a value that is protected. Then, it will discuss the primary problems associated with data mining, particularly how those problems will impact individuals when the information obtained is used by the government. Next it will describe the current landscape of privacy protections. Finally, it will examine many proposed solutions using a stakeholer/method perspective model to comparatively analyze the solutions. With that analysis, it will describe the findings and implications of that analysis.

CHAPTER 1. OVERVIEW

The Fourth Amendment has, since its adoption in 1791, protected people in the United States from unreasonable search and seizure. New technologies have exacerbated many grey areas in the law by allowing law enforcement officials new ways to gain information about the individuals they are investigating. However, perhaps the most troubling impact of technology on legal protections to privacy is third party data brokers, who sell large amounts of personal information they have gathered to law enforcement entities. This paper will explore the role of these private data mining companies in government surveillance and law enforcement proceedings and the current and potential impact they have on personal privacy.

1.1 Constitutional Protection Overview

The Constitution of the United States exists to outline the fundamental rules by which the US government will operate. One primary goal is to protect citizens from government by limiting the power of government. The Bill of Rights prohibits the government from infringing on fundamental freedoms, such as free speech and citizens' rights to due process. The Fourth amendment, which ensures

> The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath of affirmations, and particularly describing the place to be searched, and the persons or things to be seized.

The data flow process between an individual and a public entity, such as the federal government or local law enforcement, instituted in US society as desbribed by that amendment is

illustrated in Figure 1.1.

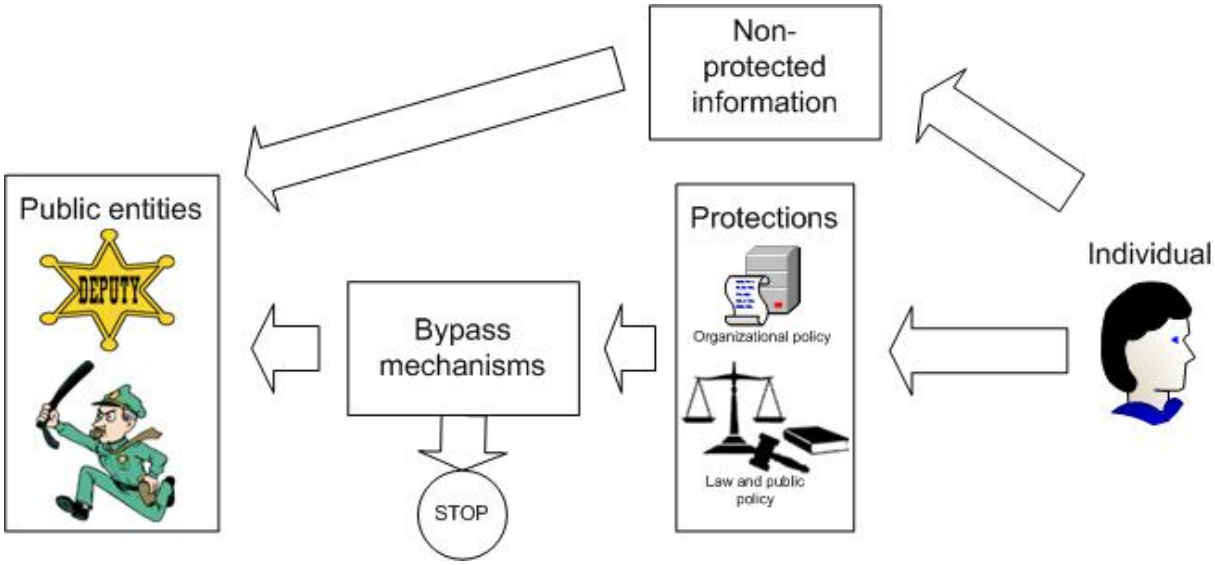

Figure 1.1 Data Flow Model

In this diagram, the arrows depict the flow of information about an individual to the public entity trying to collect information. All the information will originate from the individual, whether directly by his actions or indirectly by his iteractions with other entities. When the public entity attempts to collect the information, it must go through the following basic process:

1. Determine what protections exist for the information
2. Attempt appropriate bypass mechanisms to circumvent those protections
3. If successful, collect information

Take, for example, the situation where the local police suspect and individual committed a crime. Some informtion bout the suspect can be gathered freely because it is not protected, like asking his neighbor about the character of the individual. However, if evidence supporting the police's claim could be found inside the individual's home, the police cannot freely enter, due to Fourth amendment protection. To bypass this protection, the police can attempt to obtain a warrant from a local judge, which requires them to demonstrate why they have valid

cause to believe they could find evident there. If they are given the warrant, they are able to enter the home to collect the information they seek. If not, the flow of protected data is stopped.

The process depicted in the diagram is very simple, and taken at face value it oversimplifies the world because much of the problem lies in the first step; that is, in determining if the information is protected or not. In order for the information to be protected as defined by the court, there must be a "reasonable expectation of privacy" for the information, which is defined by a two-prong test:

- The government must contravene an individual's actual, subjective expectation of privacy
- That expectation of privacy must be reasonable, in the sense that society in general would recognize it as such

So, the inquiry into whether or not information about an individual is protected by the Fourth amendment is comprised of the following questions (1) does the individual expect privacy in this situation and (2) does society recognize that expectation as valid?

An individual can have a reasonable expectation of privacy to information outside the home. The following examples will illustrate cases involving information with some public element and the decision as to whether or not the information was considered private:

- In Katz v United States, the plaintiff argued that a conversation inside a public phone booth should not be able to be wiretapped without a warrant if the user attempts to ensure privacy. By shutting the door, the person inside isolated himself from eavesdropping, and therefore enjoyed a reasonable expectation of privacy. The Court agreed, holding that the Fourth amendment protects people, not places, and by closing the door, Katz maintained a resasonable expectation that his conversation would not be overheard.
- In Kyllo v United States, the police took pictures of a private home from public space across the street using thermal imaging equipment. From these images, they determined that the homeowner was growing marijuana plants and arrested him. The Court ruled

that because the technology was not available to the general public, the search was presumptively unreasonable.

- In California v Greenwood, garbage left on public property outside the home in order to be collected is not protected. Therefore, agents of government can search garbage without a warrant so long as it is in public space, like being on the curb.

- In Smith v Maryland. The police obtained the phone numbers dialed by Smith from his telephone company, who had installed a pen register at police request. Smith argued this was an unreasonable search. The Court stated that individuals knowingly disclose the numbers they are calling to the company so the call can be placed and so accurate billing can occur. The Court then found that an individual did not have a reasonable expectation of privacy when he willingly gave up information to a third party.

Smith v Maryland was decided in 1979, and its effects are the main source of the inquiry in this article. Although the Electronic Communications Privacy Act of 1986 regulated the use of pen registers, the third party doctrine of this ruling still remains in use. Unless otherwise protected by statute, an individual does not enjoy a reasonable expectation of privacy to any information voluntarily given to a third party.

Advances in information technology have led to an explosion in the use of personal information. Information can be collected, stored, transported, reorganized, and analyzed in ways that are cheaper, easier, and faster than were imagined in 1979. These technological changes, coupled with a lack of cohesive privacy policy, have left individuals in the US almost powerless to protect their information from agents of the government. The only legal protections awarded to individuals exist in separate and disjoint pieces of legislation guarding particular types of information and its uses. Protected classes of information include medical, financial, motor vehicle records, and communications. The absence of protection allows agents of government to obtain large amounts of personal information without experiencing any Fourth amendment requirement to supply a reason for needing it. The remainder of this paper will further explore the implications of this situation.

1.2 The Look of the New World

The improvement of information technology in recent decades has greatly altered the way in which US society treats information. Now more than ever, information is a commodity that is bought and sold, and the government is no exception. However, in order to demonstrate the way in which this change affects individuals, the types of information in question must first be defined. In this analysis, two classifications of information are used:

- Protected information: Information that is unattainable or unusable by government agents. Information can be protected in two ways:
    1. Constitutional prohibition or legal statute. Information in this category will require some type of procedure (i.e. a warrant) to legally obtain it.
    2. Infeasibility of collection. Information in this category will require some type of new process or technology to make its collection simpler.
- Unprotected information. Information that is freely able to be collected by government agents as they are interested and capable of doing so.

The commoditization of information has led to a society where individuals expect to disclose information about themselves every day. In fact, it is nearly impossible to fully participate in today's society without doing so. Stores ask for email addresses at checkout counters. Online newspapers ask viewers to disclose their location and income level. Newly purchased software programs require that the owner register them before using them. These requests have become so commonplace that individuals release information without a second thought. By simply participating in the basic advantages of society, people in the US give up their right to privacy because, except for a few exceptions (banking, medical, etc.), information held by third parties is considered to be unprotected information.

This effect is depicted in Figure 1.2, which updates the previous model of personal information flow.

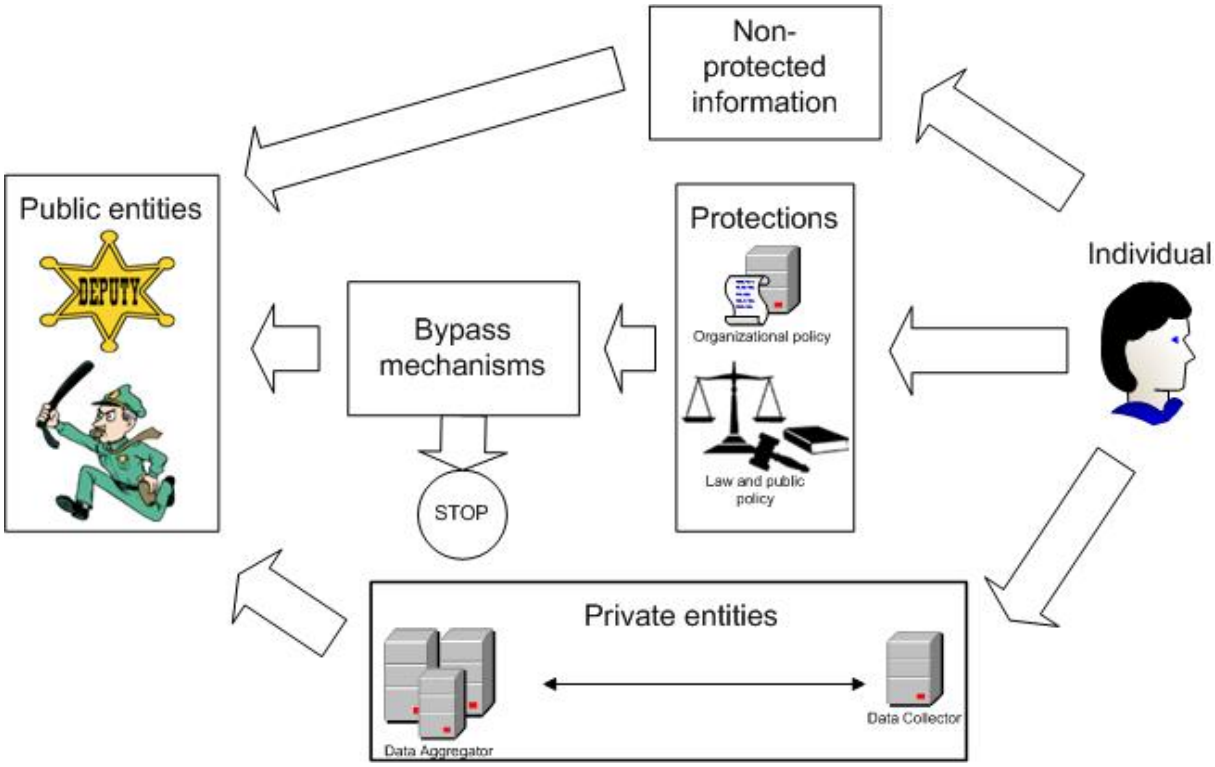

Figure 1.2 New Data Flow Model

In this image, the traditional paths for information still exist; however, a new path for information is provided - through the Private entities block. Information is supplied by the user to many third parties, aggregated, and finally given or sold to government agents.

In short, the effect of technology in light of the Smith decision is that information that was traditionally protected is rapidly moving into the unprotected category, and this greatly affects the balance of power between the individual and the government.

1.2.1 Issues in the New World

Individuals suffer several real losses in the new world. These are:

- Lack of awareness of what information about themselves is being collected
- Lack of awareness when or why informtion about them is being gathered
- Inability to correct erroneous information

- Inability to remain anonymous in daily activities
- Inability to choose which behaviors are worth the privacy risk

1.2.2 If we follow this path

Many discussions of privacy from government quickly devolve into suggestions or claims that the current system will ultimately become the Orwellian notion of Big Brother and the Thought Police. However, Solove described a different potential reality, one much more like the situation in Franz Kafka's *The Trial*. He said:

> Kafka's novel chronicles the surreal nightmare of a person who is unexpectedly informed that he is under arrest but given no reason why. A bureaucratic court maintains a dossier about him, but he has no access to this information. Throughout the rest of the novel, the protagonist desperately attempts to find out why the Court is interested in his life, but his quest is hopeless - the Court is too clandestine and labyrinthine to be fully understood[38].

It seems to be evident that our potential future is much more like that of Kafka than Orwell. Larry Hunter, a computer scientist, observed the following in 1985:

> Our revolution will not be in gathering data - don't look for TV cameras in your bedroom - but in analyzing the information that is already willingly shared[23].

Without some control over that system, we risk becoming a society where everyone knows everything about everyone else, but no one knows how they got it, what it means, or how to fix mistakes. This will not only upset the balance of power between the individual and the government - the individual will no longer have any power at all.

1.3 What to expect

This thesis will explore the impact data mining is having on privacy from agents of government. First, it will examine what exactly privacy is and why it should be examined. Then, it

will discuss theory of information, knowledge, and data mining to expose the risks inherent in today's data-driven world. Next it will survey the current landscape of privacy protection in the United States and the problems that exist today. Finally, it will analyze several proposed solutions using a two-dimensional stakeholder/method perspective model. This model will help place the solutions on a relative values scale, which will be used to draw out deeper meaning about future privacy protections.

CHAPTER 2. REVIEW OF LITERATURE

2.1 Introduction

The topic of privacy has been discussed on many different levels throughout history. It can be examined as an intrinsic philosophical concept relating to identity. It is also a societal construct denoting a place where each individual can hide from the burdens of societal life. It is a legal idea in that it provides a sphere of protection for individuals from the government. Finally, it is also a technical goal to realize via innovation. The purpose of this literature review is to examine overall attitudes and conceptions about privacy on two levels. First, it will explore what privacy actually is and why privacy is a value at all. Second, it will examine feelings towards privacy in society today.

2.2 What is privacy and why is it important?

Entire articles have been dedicated to simply attempting to define what exactly privacy means. Not only does its definition vary by perspective, say legal versus philosophical interpretations, but it may vary greatly within each school of thought. Below are a few definitions of privacy that have been articulated:

- Privacy is the right to be let alone[44]
- Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others[45]
- Privacy is a basic human right and "no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and

reputation. Everyone has the right to the protection of the law against such interference or attacks.[41]"

- In political philosophy, the private world of introspection and the private pursuit of economic interests versus the public world of action and speech[1]

What all of these concepts have in common is that privacy is something that is inherent to each individual. It is a concept that allows us as people to form our own opinions and choose whether or not to share them. It is a value that implicates the utmost sanctity of our individual autonomy, and it is a value that is instrumental in a society that allows its citizens to be individuals. Even though privacy cannot be defined, it is generally understood that if individuals lose all privacy, they will lose some characteristic of their humanity.

2.3 How does society view privacy?

Most commonly, privacy is viewed as a social construction that keeps people and government from bothering those who would prefer to be left alone. To be private is to be secluded.

First, there are a few sources that argue that privacy will become an antiquated concept, and the current uproar about it is simply the unwillingness of society to embrace change. The most notable author in this regard is David Brin, who proposed a transparent society where everyone is watched, but everyone is also a watcher [3]. He argues that today's world of secrets and realms of solitude gives space for people of 'bad intent' to do harm, whereas as it becomes more transparent, the society will also become more enlightened. He argues that individuals will feel embarassments because everything people do will be in the open, but because everyone will be vulnerable and watched the social benefits by far outweigh the small personal sacrifice.

Another common construct is the view that privacy is about a balance of power between the government and the individual[21, 30]. When discussing privacy in the societal and legal context, such as in this paper, this is the most useful construction. It demonstrates that privacy is a value, but not an absolute one. It competes against other valid interests. Just as participating in society as an individual usually causes some loss in privacy due to a loss of isolation, empowering the government to accomplish the task of policing security requires that

they be able to intrude on the private space to enforce those rules if they need to. The term 'balance' is very appropriate because it invokes important mental imagery - a scale. In order to function as a cohesive society comprised of autonomous beings, a balance must be struck between the powers of the government and the rights of the individual.

CHAPTER 3. Information and Data Mining

3.1 Information Theory

3.1.1 Data, Information, Knowledge and Wisdom

To start, it is important for the purposes of this inquiry to understand what can be learned about a person and specifically what a certain level of comprehension implies. To accomplish this, detailed definitions of data, information, knowledge, and wisdom will be used to be clear about what exactly is being described[2, 5, 7].

- Data: a fact without further relation to anything else, therefore having no meaning or value because of lack of context and interpretation.
- Information: data that is somehow related to something else to be useful or provide meaning.
- Knowledge: application of organized information to infer or predict something by recognizing patterns
- Wisdom: deeper understanding of the knowledge, grasping principles rather than patterns

This structure is called the DIKW hierarchy. Basically, it describes a system of different levels of understanding. Some descriptions relate DIKW to a pyramid, like Maslov's hierarchy of needs where one level must build upon another.

The breakdown of the characterization of how facts and learning relate will be vital in interpreting the topic of privacy in the context of the Information Age. Since this paper focuses on the uses of data and information to gain knowledge and wisdom by government

agents, the example below is set in a law enforcement scenario.

Data: A woman bought a gun.

Information: The woman who bought the gun very recently discovered that her husband was unfaithful.

Knowledge: Spouses are often very upset after discovering such news and may be prone to violence against the spouse. Therefore, the woman may have bought the gun to shoot her husband.

Wisdom: Spouses tend to be upset in this situation because they view infidelity as a violation of the sacred promise that they made to each other in getting married.

As it may be obvious, only the level of knowledge is really vital for law enforcement to do its job in this scenario. Once the pattern recognition occurs, the police are likely able to provide the adequate protection. In the data mining context, the level of wisdom and knowledge will be combined to one level that signifies deeper understanding of the situation, whether it be principles or patterns, due to analysis of the information available.

3.1.2 Can a computer *know* anything?

In a discussion about email privacy in the law of imputed knowledge, a researcher argued that electronic records should be considered part of the knowledge of the organization that can access them instantly[4]. This law states that knowledge of an organization includes all the knowledge of the principles of that organization, not just the knowledge that the principles share with the organization. This is because the information is "ready-to-hand," or could be shared with that organization at a moment's notice. This same principle applies to computer records, because they can be searched almost instantly to provide the information that they hold, even if the organization is not aware of it. However, this would not directly apply to paper records, because a human must know they contents for the organization to be aware thay they are there. Therefore, a computer does not really know anything, but the organization that can instantly and meaningfully access the data contained by that computer 'knows' all its data.

3.2 Data Mining

Data mining is defined as the process of extracting hidden patterns from data. Put in the DIKW context, the purpose of data mining is using data to gain knowledge, as shown in Figure 3.1.

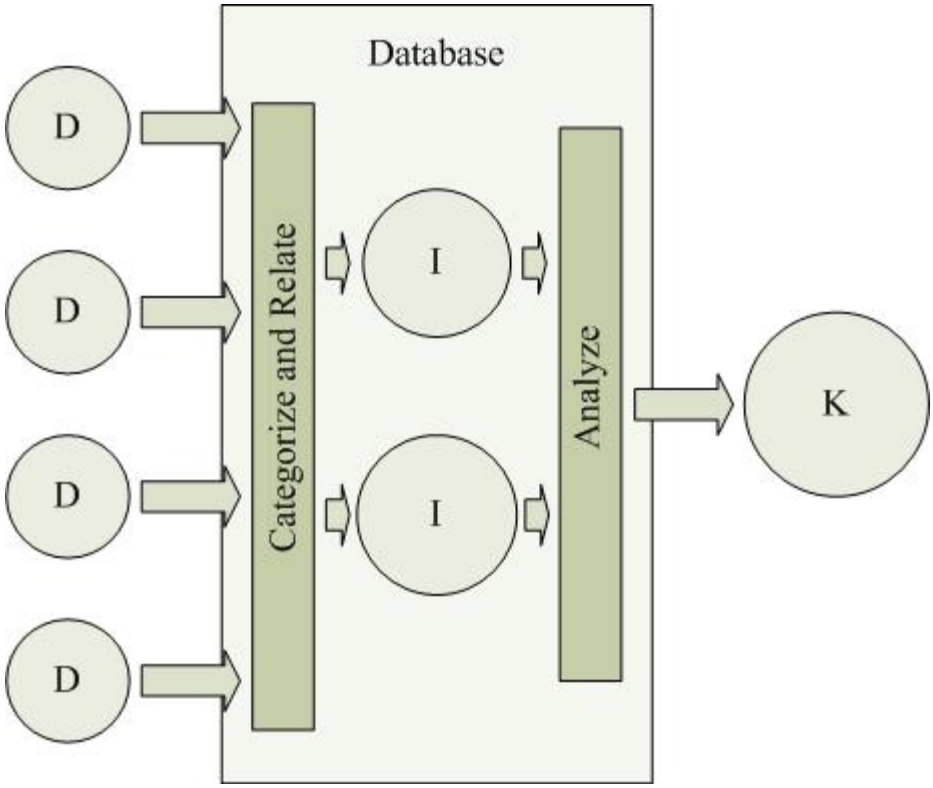

Figure 3.1 DIKW model in Data Mining

Government data mining is commonly used in two ways: discovering a new pattern or predicting future behavior based on past[40]. Knowledge discovery identifies patterns found in the collected data to reveal some piece of additional information about an individual, a group, or society. This is used commonly in research and business contexts to discover patterns, such as in medical trials of new drugs or the effectiveness of marketing campaigns. Prediction analyzes current information to infer something about the past, present, or future. This has been used in government programs to predict likely terrorists, such as the Total Information Awareness program by the CIA[**?**].

The data mining process consists of five major elements[27]:

- Acquire data and transform it to the appropriate format for the database
- Store and manage the data
- Provide data access to analysts
- Analyze the data using software
- Present the data and analysis in a useful format

These five steps let entities transform raw data about people to meaningful patterns. For example, grocery stores often mine point of sale (POS) information to increase their profitability. For example, they track what days, times, and in what combinations products are purchased by customers. They may use this information to track which displays are most effective at marketing, which items should be placed near each other because they are likely to be purchased together, or on what days certain items should be full price because they are more frequently purchased at that time.

Data mining obviously cannot discover patterns which do not exist, nor can it find patterns from data that has not been collected. Nevertheless, the effect of data mining is increasing exponentially in today's society. The amount of digital information in existence doubled in the three years before 2003[15]. In addition, innovations leading to improved processing power, data storage, data capture, and analysis are increasing the usefulness and effectiveness of data mining.

Christopher Slobogin, a law professor at Vanderbilt University, identified three basic modes of data mining performed by law enforcement[35]:

1. Target-driven data mining, which is also known as subject-based data mining. This is a search to obtain information about a particular person
2. Match-driven data mining. This is a search to determine if a particular person has already been identified as a person of interest. An example of this is when an airline passenger's identity is compared to those names on the no-fly list to determine his eligibility to fly.

3. Event-driven data mining, which is also known as pattern-based surveillance. This is a search to try to discover the perpetrator of a past or future event.

3.2.1 Data Mining in Society

In today's society, individuals create multitudes of data every day. The following entities are examples of those that save data for typical daily activities:

- Cell phone companies log every call make, its length, and probably the GPS location of the phone when it was placed
- ISPs log every web site visited
- Email providers log email messages, including sender and recipient
- Credit card companies record any charges made
- Libraries, movie rental places, etc. record rental history
- Banks record many business transactions, like ATM withdrawals and checks
- Utility companies monitor electricity and gas usage

In addition, public records describe many of the interactions between individuals and society. For example,

- Births, deaths, and marriages
- Criminal and civil court proceedings
- Motor vehicle registration
- Property ownership records

The knowledge that can be obtained from the aggregate of all the information that one person will produce, particularly over time, can be very revealing. Not only the person's associations and business dealings, but also religious and political affiliations, interests, health, finances, lifestyle, and psychology may be ascertainable from intensive data mining. A person may feel

less exposed purchasing a self-help book from Amazon because other people may not witness them doing so in the moment, but that individual may not realize that not only does Amazon remember that purchase, but it may sell the record of the purchase to marketers who target people with self-help interests.

However, the issue of privacy from business and data mining for marketing has been explored by many other people and is not the issue here. Instead, the focus is the impact this practice has had on the government's use of private information.

3.2.2 Data Mining Issues

Data mining methods can greatly increase the ability of law enforcement to prevent and prosecute crime, whether it is fraud or terrorism. However, each of the methods are prone to the following challenges.

3.2.2.1 Analytical errors

Three problems lead to errors in the analysis of data mining.

- Data inaccuracy. Data mining often incorporates data from multiple sources, the accuracy of which is not always verified. Incorrect data can have huge impacts on the analysis of the information. For example, credit reports are frequently used to determine eligibility for financial credit for loans or credit cards. However, in 2004 approximately 1 in 4 credit reports were found to contain errors serious enough to result in a denial of credit, employment, or housing[6]. These errors occur in a variety of ways, from the mis-entering of information, mistaken identity, or even identity theft.

- Incomplete information. Data mining gains knowledge by analyzing information, but it often does so with only a subset of the total data that describes the situation, which could lead to misinterpretation. This problem occurs in making judgments in life all the time. In data mining, incomplete information could lead to a distorted analysis by a computer or an analyst.

- Difficulty creating accurate profiles and effective algorithms. This is a problem that primarily affects event-driven data mining, which attempts to determine likely candidates for the perpetrators of past or even future crimes. In order to accomplish this goal, the analyst must determine what the distinguishing characteristics of individuals who would commit crimes and only those individuals.

Figure 3.2 depicts how a target-driven data mining operation proceeds from the standpoint of a piece of data. The raw data is matched to an identity (ie, correlated to a subject) and stored as information. The piece of information about the subject in question is then combined with other information supposedly about that same subject. The result is the ability to determine whether or not the individual meets the criteria.

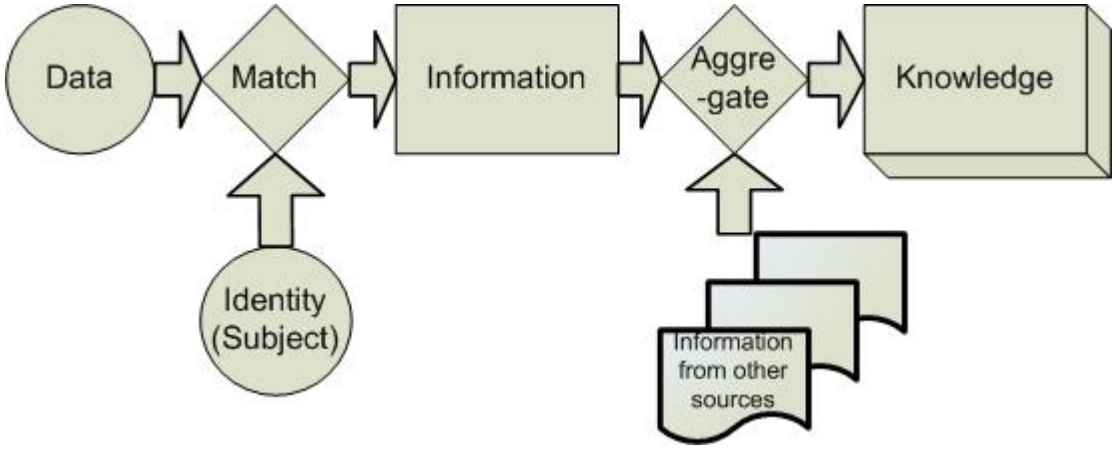

Figure 3.2 Target Driven Data Flow

This type of process occurs every day across the United States, such as when a credit report is computed. The credit report contains data mined from various sources, like bank account information and credit uses and abuses. This process is actually very difficult to complete without error. Each step has some degree of uncertainty, and therefore introduces room for error. Figure 3.3 depicts some of the questions that become apparent at each step.

The three main problems described at the beginning of this section could cause an undesirable result at the end of this mining process:

- Data inaccuracy: The wrong identity could be matched to data, such as when identity

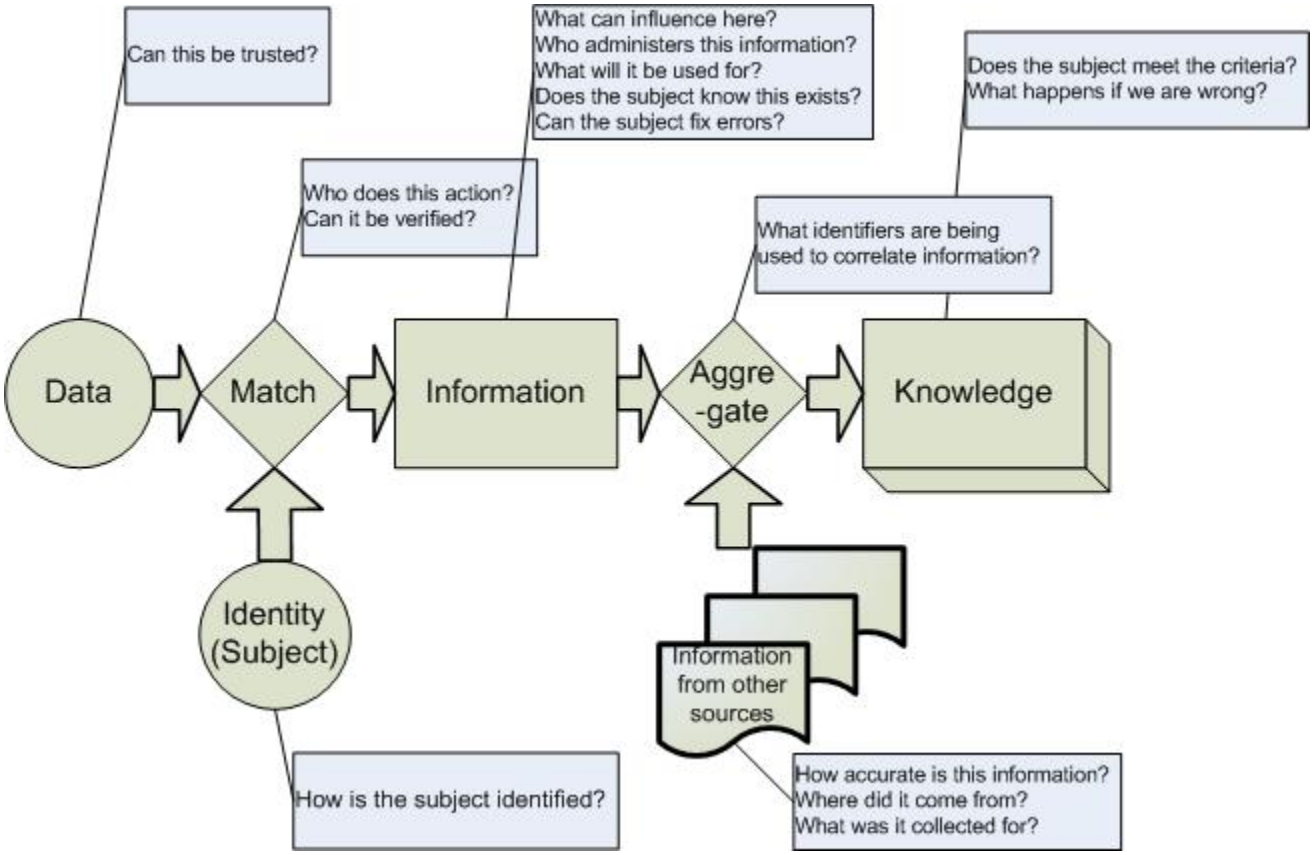

Figure 3.3 Target Driven Data Flow with Comments

thieves open credit cards using someone else's name. The individual will be associated with reckless spending he did not do, which could harm his ability to get credit.

- Incomplete information: If a match was done incorrectly, this person could have some outstanding debt that was not attributed to him because the identity information was incorrect. Therefore, the bank may be unaware that this person has a history of financial delinquency, therefore decreasing their desirability as a loan customer.

- Difficulty creating accurate profiles and effective algorithms: In order to maximize profits and minimize losses, the bank is interested in developing a profile that will allow them to select customers who will be able to pay the loan on time and give them the maximum amount of money that they can afford. This requires the bank to develop profiles to determine who those customers are and how much money that is in their cases. Getting

this process wrong could mean that the bank will lose money on unreliable customers or simply that they could have earned more money by loaning a larger amount.

3.2.2.2 Inference Problem

In some data collection systems, the desired knowledge is the large-scale trends, not the individual ones. Therefore, these systems may anonymize the data in order to protect privacy. However, it is sometimes possible to infer private information from anonymized aggregate data. This is the inference problem[8]. Take, for example, a politician who polls registered voters to determine who is likely to vote for him and who is not. He may group the voters into categories of race, gender, wealth, profession, or location, in order to try and determine trends. However, when analyzing the results, it may be possible to identify the only African-American woman engineer making between 50,000 and 70,000 annually living in Ames, Iowa. In this instance, by collecting anonymous information about voters' preferences and classifications on a large scale, the researcher can identify the vote of a single person. The privacy of one's vote is a right that is historically well-protected in American society, but data mining can potentially significantly reduce or even remove the privacy of that choice. This is a serious problem because studies have found certain combinations of data to be particularly revealing. One study determined that 87% of the US population could be uniquely identified with only his 5-digit zip code, birthdate, and gender[43].

3.2.2.3 Base Rate Fallacy

Another problem, which is particularly important for the use of data mining in legal contexts is known as the base rate fallacy. This refers to an issue that is found in the application of detection theory to data mining. Detection theory relates to the ability to distinguish signal from noise, for example, when detecting fraudulent use of credit cards in order to reduce the amount of money they lose due to the illegal activity, since most credit card companies do not hold the card owner to be liable for the costs. Therefore, they try to automatically detect when a credit card is being illegally used. The classifications of card use are shown in Table

3.1:

Table 3.1 Possible Detection Outcomes

		Is the card being used fradulently?	
		Yes	**No**
Does the system detect the use as fraud?	**Yes**	True Positive	False Positive
	No	False negative	True negative

Ideally, the system would correctly identify use in every instance, therefore only outputting true positives and true negatives. However, this is impossible to implement in almost all cases. Obviously, the credit card company is concerned about false negatives, because this means they miss the fraudulent activity and must pay for the costs of the illegal purchases. However, the credit card company is also worried about false positives because often the credit cards are deactivated when fraud is detected. If this happens often to customers while they are using their own cards, they may be likely to change companies because of the recurring inconvenience of this fraud protection.

Often, balancing false negative and false positive rates is a tradeoff - reducing one means increasing the other. Therefore, entities that use detection techniques must determine acceptable thresholds for these rates.

Data mining to detect credit card fraud works well because credit card fraud is a pervasive problem that occurs every day. In addition, it is easy for detection algorithms to generate an accurate profile describing fraudulent behavior because so much fraud exists to be studied. However, terrorist detection systems that work like the one described here are not always effective. This problem occurs when the base rate, or number of true positives as a proportion of the whole sample space, is too low. For example, detecting terrorist activity suffers from the base rate fallacy. Terrorist plots occur so rarely and usually do not follow typical patterns, rendering data mining for patterns related to terrorism ineffective. Bruce Schneier illustrated this problem with the following hypothetical figures:

> Let's look at some numbers. We'll be optimistic – we'll assume the system has a one in 100 false-positive rate (99 percent accurate), and a one in 1,000 false-

> negative rate (99.9 percent accurate). Assume 1 trillion possible indicators to sift through: that's about 10 events – e-mails, phone calls, purchases, web destinations, whatever – per person in the United States per day. Also assume that 10 of them are actually terrorists plotting.
>
> This unrealistically accurate system will generate 1 billion false alarms for every real terrorist plot it uncovers. Every day of every year, the police will have to investigate 27 million potential plots in order to find the one real terrorist plot per month. Raise that false-positive accuracy to an absurd 99.9999 percent and you're still chasing 2,750 false alarms per day – but that will inevitably raise your false negatives, and you're going to miss some of those 10 real plots[33].

3.2.2.4 Lack of individual control

The questions in Figure 3.3 implicate more than just concerns for the bank in question in ensuring that all qualified and zero unqualified customers get credit. The individual seeking credit has a stake in the outcome as well. Despite the fact that the process is entirely about the individual, they have no control.

A person's information could be data mined for many different purposes, from credit analysis to marketing to terrorism detection. It can be used, shared, repackaged, manipulated, and sold in an infinite number of ways. This leads people to ask:

- What data about them is being collected?
- Where does it come from? Is it even originated by them?
- How do they find out what information about them is out there?
- How can you correct any incorrect information?
- If data is collected for one legitimate purpose, what else will it be used for?
- Are there ways to reduce the generation and sharing of this information?

The upcoming chapters will explore further the rights of the individual regarding his information and evaluating the various solutions for guaranteeing those rights.

3.3 Government and Personal Information

It has been mentioned in several scenarios in this section that government agents try to collect information to pretect the public welfare, such as detecting and stopping fraud and preventing terrorism. Figure 3.1 depicted the application of the DIKW model in information gathering techniques such as crime detection. In the introduction, I proposed two categories of information in the law enforcement context:

- Protected information. Information that is unattainable or unusable by government agents. Information can be protected in 2 ways:
    1. Constitutional prohibition or legal statute. Information in this category will require some type of procedure (i.e. a warrant) to legally obtain it.
    2. Infeasibility of collection. Information in this category will require some type of new process or technology to make its collection simpler.
- Unprotected information. Information that is freely able to be collected by government agents as they are interested and capable of doing so.

Historically, a large amount of information was protected by the second method: infeasibility of collection. Tracking someone's movements was difficult because someone had to follow them, requiring a lot of time. Determining who a person corresponded with over the past six months could only happen by asking the subject or finding every person that they corresponded with - the post office does not log who sends letters to whom. Finding what bank someone used in order to obtain his financial information required that an investigator contact every bank in the area and ask if that individual is a customer. Now, all these pieces of information are logged routinely and stored digitally. They can be collected, shared, and searched almost instantly. Not only is it feasible to collect it, it is much easier to gather information by data mining than using other traditional methods of investigation.

The following chapter will go through privacy protections in law and practice in order to determine exactly what types of information are protected. In doing so, it will demonstrate how much information today is unprotected from use by government agents and what types of problems this can cause for the individual.

CHAPTER 4. Current Privacy Protection

Individual protections in the United States come from a variety of sources. First, the Constitution enumerates those rights which are viewed to be fundamental in our society, and it serves to establish the balance of power between government and the individual. Although those rights are not absolute, any interference with those rights requires a compelling government interest. However, the Constitution only protects those rights which are considered to be fundamental and enduring. Federal and state laws are enacted to enhance those protections and provide additional guidance as to the way government and individuals should act.

Treatment of privacy varies greatly across the world. From very strict control of personal information throughout the European Union to the extensive government intervention in life in China, expectations of privacy are very much defined by the society in question.

In addition to legal and social frameworks, some technical measures for privacy protection exist. These solutions can be implemented at the personal, organizational, or even national level to provide privacy protection.

This chapter will explore each of these areas in greater detail.

4.1 Privacy and the Constitution

The Constitution is the foundation for the US system of government and it establishes the law of the land. Its purpose was not only to establish a working government, but to do so in a way that would assure the essential liberties of all the people within its scope, as seen in the beginning of the Preamble:

> We the People of the United States, in Order to form a more perfect Union, establish Justice, ensure domestic Tranquility, provide for the common defence, promote the

> general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America

This section will explore to what extent, if any, the "Blessings of Liberty" clause encompasses the protection of privacy.

4.1.1 Is privacy in the Constitution?

The word privacy does not appear anywhere in the text of the Constitution of the United States. However, this does not mean that privacy is a new value nor a right that is not protected. In perhaps the most influential law review in modern history, Brandeis asserted that the "right to life has come to mean the right to enjoy life - the right to be left alone[44]." In 1890, he was advocating privacy as a fundamental individual right by explaining how a concept in the text encompassed or implied that right.

The job of the Supreme Court throughout history has been to interpret the meaning of the Constitution as it is applies to certain situations, and in doing so they have found a right to privacy implied by the text. The first mention of modern privacy came in the ruling of NAACP v. Alabama. The Court ruled that the NAACP should not have to disclose its membership list to the state of Alabama because the First amendment protects the "freedom to associate and privacy in one's associations." Just seven years later, the Court found that privacy was a "penumbral right" in Griswold v. Connecticut. The court reasoned that although privacy is not specifically mentioned in the Constitution, it is implicated in the First (free speech), Third (protection from quartering soldiers), Fourth (search and seizure), and Fifth (self-incrimination) amendments. Furthermore, the Ninth amendment allowed the Court the flexibility to discover new rights as belonging to the people. In finding this right to privacy in Griswold, the Court invalidated a law in Connecticut that banned contraceptives, saying that the Constitution safeguarded individual automony for decisions that involve their bodies and family.

4.1.2 Judicial interpretations of privacy since Griswold

Since Griswold v. Connecticut, the right to privacy has become a standard facet of jurisprudence in the United States. However, its application and interpretation have been often remolded with subsequent decisions. This section will highlight judicial interpretations of privacy in order to try to provide the context of today's understanding of this right.

Whalen v Roe In Whalen v Roe, the Court now recognized that the "zone of privacy" protected by the Constitution was realized in two main forms:

1. Decisional privacy, or "independence in making certain kinds of important decisions"

2. Informational privacy, or "individual interest in avoiding disclosure of personal matters"

However, the Court has not really expounded on the informational privacy right any further, although it has been used in many circuit courts[37].

Katz v United States Katz was the beginning of a revolution in Fourth amendment law. The Court determined the wiretapping of a public phone booth, where Katz was making a call, was a violation of the Fourth amendment. In its opinion, the Court said:

> What a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection. But what he seeks to preserve as private, even in an area accessible to the public, may be consitutionally protected.

In his famous concurrence in this case, Justin Harlan established the "reasonable expectation of privacy" test. This test is comprised of two inquiries:

1. Does the person exhibit an actual or subjective expectation of privacy?

2. Is that expectation one that society is prepared to recognize as "reasonable"?

This test has been the fundamental basis for determining Fourth Amendment protection since 1967.

US v Miller and Smith v Maryland These two cases established the third party doctrine of modern privacy law. In each case the court held that financial records (Miller) or pen registers (Smith) were not private information because it was willingly shared with a third party. They reasoned that sharing information eroded Constitutional protection because the individual, by exposing the information to others, could no longer have an expectation of privacy. Although later the Electronic Communications Privacy Act of 1986 regulated the use of pen registers, the third party doctrine established in these rulings still remains in use. Unless otherwise protected by statute, any information voluntarily given to a third party does not enjoy a reasonable expectation of privacy.

Kyllo v US The police took pictures of a private home from the public space outside using thermal imaging equipment. From these images, they determined that the homeowner was growing marijuana plants and arrested him. The Court ruled that because the technology was not available to the general public, the search was presumptively unreasonable.

4.2 Privacy Protection by the Legislature

The interpretation of the Fourth Amendment by the Court allows the government much freedom to gather information about individuals. The increasing digitization and sharing of information is leading to a phenomenal amount of information being available about every person who chooses to participate in US society. The various legislative bodies in the US have recognized this and taken action to provide additional protections in response. This section will describe many of the statutes that have been enacted to protect privacy.

4.2.1 Federal Statutes

Federal law does not holistically protect privacy, but instead does so as a patchwork of laws protecting very specific interests. This section will describe the current privacy protections and government powers regarding information.

4.2.1.1 Financial Statutes

Various statutes have been enacted surrounding financial information.

Fair Credit Reporting Act (FCRA) - 1970 This act regulates consumer credit reporting agencies. It requires these agencies to make their records available to the subjects of the records and allow individuals to correct information about themselves and regulates disclosure of the information.

Bank Secrecy Act - 1970 This act compelled banks to start keeping records of certain types of transactions to aid in detecting fraud and money laundering.

Right to Financial Privacy Act - 1978 This act was created to restrict access to the records that were now kept due to the Bank Secrecy Act. In order to get bank records, the government must have a court order or other formal request.

Gramm-Leach-Bliley Act - 1999 This act requires financial institutions to notify customers of their privacy policy. If the institution wants to disclose any nonpublic personal information to nonaffiliated third parties, they must first notify the customer and allow them the opportunity to opt-out. It also outlawed a practice known as pretexting, which is defined as obtaining financial information about another person under the pretext of being that individual.

4.2.1.2 Communications Statutes

Various statutes have been enacted regarding communications.

Omnibus Crime and Control Act of 1968 Also known as the Wiretap Act, it extended wiretap regulations to state officials and private actors. This established that a warrant would be required, except for very specific circumstances, to obtain a wiretap.

Foreign Intelligence Surveillance Act (FISA) - 1978 This act established a separate legal process for electronic surveillance being used to gather foreign intelligence. The standards for obtaining probably cause are perhaps lesser than those of the Wiretap Act, but were greater than the previous unchecked gathering of evidence for this purpose.

Electronic Communications Privacy Act (ECPA) - 1986 This act amended the Wiretap Act to extend coverage to new forms of communication including cellular phones, email, computer transmissions, and pagers. The ECPA restricts both the interception of communications as well as the searching of stored communications (known as the Stored Communications Act). A third component also limited the use of pen registers and trap and trace devices.

Privacy Act of 1974 The Privacy Act was passed in response to a report done by the US Department of Health Education and Welfare, titled "Records, Computers, and the Rights of Citizens." The report recommended a set of Fair Information Practices to stop the problems it observed, namely

> An individual must increasingly give information about hiself to large and relatively faceless institutions, for handling and use by strangers - unknown, unseen, and, all too frequently, unresponsive. Sometimes the individual does not even known that an organization maintains a record about him. Often he may not see it, much less contest its accuracy, control its dissemination, or challenge its use by others[42].

The act responded to these concerns by regulating the collection and use of records by federal agencies and allows individuals the rights to access and correct their personal information. This Act is specifically targeted at restricting the ability of the federal government to amass large databases of information about citizens.

4.2.1.3 Other Acts

Several other types of information are protected via individual statutes. A few examples are listed here.

Video Privacy Protection Act of 1988 Enacted in response to the uproar after Supreme Court Justice nominee's Robert Bork confirmation hearings, where people obtained and used his video rental history as evidence to his character, this act forbids disclosure of video rental or purchase information[36].

Driver's Privacy Protection Act of 1994 This act was the first and only direct regulation of the states' use of data at the federal level. It required that states must first obtain a person's consent prior to disclosing his motor vehicle record information. The practice of selling those records had generated a large amount of revenue for many states prior to this act.

Health Insurance Portability and Accountability Act of 1996 (HIPAA) This act required that an individual must authorize the use and disclosure of his health information when not used for treatment, payment, or health care operation. However, not all medical records are covered. HIPAA only applies to records maintained by health plans, health care clearinghouses, and health care providers. In addition, HIPAA provisions only require that law enforcement have a subpoena, rather than a warrant, to obtain the information.

Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act (USA PATRIOT Act) The Patriot Act, passed after September 11, 2001, updated the ECPA and FISA. It broadened the pen registers regulation to now include the addressing information on emails and IP addresses. It increased the types of subscriber records that could be obtained from providers of communication services, like ISPs. It also broadened the purpose for which wiretaps under FISA could be conducted, now only requiring that the wiretap be used in investigations where foreign intelligence gathering is "a significant purpose."

4.2.2 Privacy in the States

The Driver's Privacy Protection Act is a rare federal statute, at least at the current time, because it regulates the ways in which the states can use information. Most other aspects of state recordkeeping are up to the states to regulate themselves if they so choose.

The Electronic Privacy Information Center (EPIC) website displays a state by state list of different privacy areas and if those states protect information in that area. Table 4.1 contains the national data and data about five states: California, Florida, Iowa, Connecticut, and Texax[28]. A "Y" means that the state has some law regulating that type of information. An "N" means no such law exists.

Table 4.1 State Privacy Laws

		State				
Category	US	CA	FL	IA	CT	TX
Arrest Records	N	Y	Y	N	Y	N
Bank Records	Y	Y	Y	Y	Y	N
Cable TV	Y	Y	N	N	Y	N
Computer Crime	Y	Y	Y	Y	Y	Y
Credit	Y	Y	Y	Y	Y	Y
Criminal Justice	Y	Y	Y	Y	Y	N
Government data banks	Y	Y	Y	Y	Y	N
Employment	Y	Y	Y	Y	Y	N
Insurance	Y	Y	Y	N	Y	N
Mailing Lists	Y	Y	Y	Y	Y	N
Medical	Y	Y	Y	Y	Y	Y
Miscellaneous	Y	Y	Y	N	Y	N
Polygraph Results	Y	Y	N	Y	Y	Y
Privacy Statutes	Y	Y	Y	N	N	Y
Privileges	N	N	N	N	Y	Y
School Records	Y	Y	Y	Y	Y	Y
Social Security Numbers	N	N	N	N	N	N
Tax Records	Y	N	N	N	N	N
Telephone Solicitation	Y	Y	Y	Y	Y	Y
Testing	N	N	Y	Y	Y	N
Wiretaps	Y	Y	Y	Y	Y	Y
Cited from Epic.org						

4.3 International Privacy Law

The notion of privacy, and whether or not it warrants protection, differs internationally. The European Union seems to be at the forefront of privacy protection. The United States seems to fall in the middle of the spectrum, allowing more commercial use than many and less government use than some. This section will summarize a few international regimes, focusing primarily on the EU, whose protections are slowly influencing the rest of the international data economy.

International privacy law began with the Organization of Economic Cooperation and Development (OECD) Privacy Guidelines, which where largely based on the Fair Information Practices that HEW expressed in the 1973 report. The eight guidelines are[24]:

1. Collection limitation
2. Data quality
3. Purpose specification
4. Use limitation
5. Security safeguards
6. Openness principle
7. Individual participation
8. Accountability

The Privacy Guidelines influenced much of the privacy policy adopted by the European Union.

4.3.1 Asia-Pacific countries

The more developed nations in this region, such as Australia, New Zealand, Hong Kong, Korea, and Japan have fairly sophisticated privacy standards. Many of them follow a model similar to the US, featuring heavier regulation of government use of information than commercial use. However, other Asian-Pacific countries lack much privacy protection. For example,

Singapore currently only uses voluntary, self-regulatory schemes. The People's Republic of China has only adopted indirect protections of privacy, and there is little sign that they intend to reform to meet EU standards[43].

4.3.2 European Union

The first directive on data privacy (Directive 95/46/EC on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data - also known as the EU Directive), is a comprehensive law that protects personal information maintained by a broad range of entities. Joe Reidenberg said

> The background and underlying philosophy of the European Union Directive differs in important ways from that of the United States... The United States has, in recent years, left the protections of privacy to markets rather than law. In contrast, Europe treats privacy as a political imperative anchored in fundamental human rights[31].

It is very influential internationally because it prohibits the transfer of personal data to countries that do not provide adequate levels of privacy protection.

4.4 Technical Means for Privacy Protection

Three basic categories of technical solutions have been proposed to help protect privacy: identity hiding, information hiding, and information minimization. Each of these will be discussed in the following subsections.

4.4.1 Identity Hiding

Various techniques for maintaining privacy involve some form of anonymity or pseudonym as a way to hide the identity of the subject. These methods protect privacy by making it hard for anyone with the data to connect it to an actual person. Some examples of existing identity hiding technologies or methods are:

- Tor, which is an anonymizing routing technology
- Proxies, which allow anonymous web surfing
- Anonymous remailers, which hide where email originated
- Avoiding registered accounts
- Deleting or not accepting cookies, which are often used to track user actions
- Using a NAT, which hides IP addresses
- Data aggregation, which allows information to be collected without identifying individuals
- Obfuscation or chaffing, which entails adding random or misleading information to conceal or obscure actual behavior

Tor, proxies, remailers are ways for a user to attempt to remain anonymous in web surfing and other online activities. In fact, Tor was developed by and is used by the government for use in intelligence gathering, in instances where the web site it visits must not know that it was visited by a government agency[46]. These technologies keep the identity of the user secret from the entity it visits, as well as anyone else watching, by not revealing the location where the traffic originated However, this does not prevent the users from identifying themselves by the content they send.

Data aggregation and obfuscation are commonly used by people who mine data to try and keep the anonymity of the data contributors. This is common practice in research and other areas where the identity of single users is not important for the analysis.

As described in the previous section in the discussion of the inference problem, it does not require a name or a social security number to uniquely identify a person. Figure 4.1, which was created by Greg Conti in *Googling Security*, shows a hypothetical diagram depicting the ways in which a subject's activity is correlated[5]:

Basic identity hiding techniques can help stop this from occuring, or at least limit its effects. However, many of the services available online require a user to identify himself in

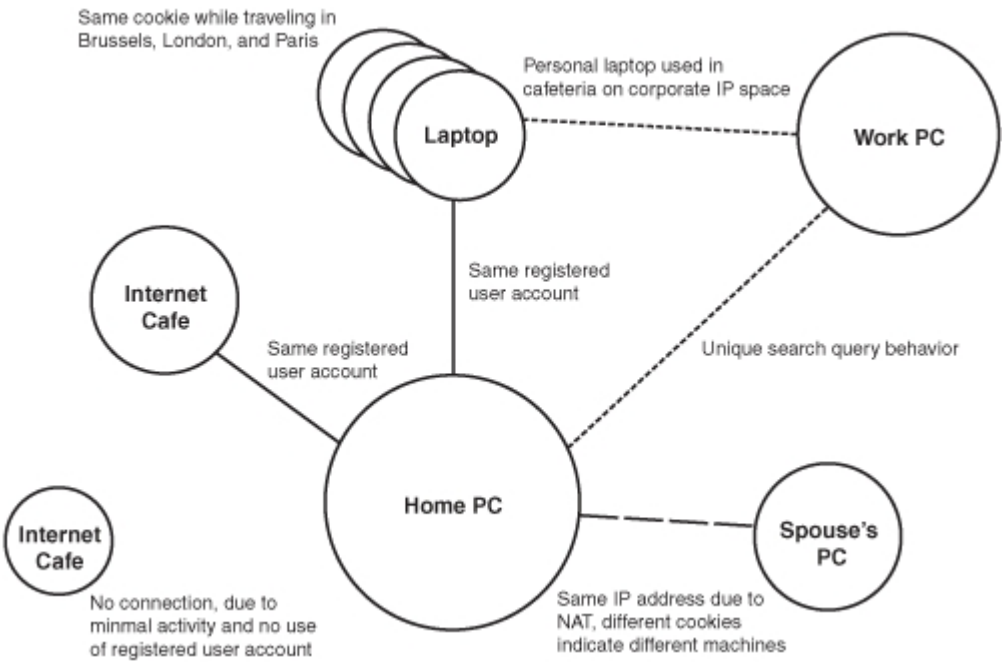

Figure 4.1 Correlating User Activity, by G. Conti

order to participate. Meaningfully using email, online banking, and various other services without allowing the activity to be linked to the same individual would take a great deal of individual effort and still be likely to fail.

4.4.2 Information Hiding

Information hiding refers to ways to conceal information. Techniques for accomplishing this are:

- Encryption, which will change the content to a (presumably) unreadable format, except to people who know or can break the code. Encryption can be used for most types of information, like emails or the information stored on a personal computer
- Hidden partitions, which are a way to attempt to create a secret location on a computer for information that an individual would prefer not be discovered
- Refraining from accessing personal information in public places. For example, do not check personal email or banking at work. The workplace has the authority to watch

whatever its employees are doing online and could be logging the traffic.

4.4.3 Information Minimization

This technique involves only generating information that is necessary, as well as removing it once it is no longer needed. This could include:

- Content filtering during web browsing and disallowing third party content such as banner ads, referrer values, and web bugs
- Deleting cookies frequently
- Deleting browsing history
- Performing complete deletes, or "zeroing" digital information once it is no longer needed

4.5 Problems with the Current Protections

Right now, technology and practice in collecting and sharing information is by far outpacing technology and law to police privacy. While the increased availability of information can have great benefits to individuals, to companies, and to government, personal privacy is still a legitimate concern, and one in danger of being extinguished if this trend continues unchecked.

4.5.1 Why can the market not take care of privacy?

It is possible that technical advancement and market forces will work in a way that will favor privacy, just as it is possible that those forces will continue to act to demolish privacy. If a large number of consumers demanded privacy and were willing to pay for it, the market would presumably solve the problem. However, users at the moment like their online services to be free, even at the cost of some privacy, and marketing companies are willing to pay for those services in return for the information. Until that paradigm changes, it is not likely that the lack of privacy protection will be remedied without government intervention. Targeted marketing provides strong incentive for companies to ensure the current system stays in place.

4.5.2 Why can technology not take care of privacy?

Technology seems to always be a cycle, never an end. Not long after new protections are created, someone finds a way to beat them and gain access. Then another protection will be developed, and so on. Some information will always need to be shared and ultimately be associated with real people; society in the forseeable future will demand continued interconnectedness. Technology will play an important role, both in the protection and the destruction, of privacy - but it will never be a final solution.

4.5.3 So it is up to the government

The need for privacy protection is more than just a fear that the government will turn into the infamous Big Brother, but rather it is about the balance of power between the rights of the individual and the government. Although enforcement of the law and protecting the safety of the citizens are important goals, those aims do not automatically trump the rights of the individual citizen. To protect this balance, the Constitution requires that there be a *reason* for the government to search or seize the papers or effects of an individual.

However, that statement has several ambiguities which must be addressed in its application. The new role that digital information is playing in society and the ways in which it is different from traditional physical evidence have only exacerbated the problem. The important questions that must be answered are:

1. What does it mean to search digital information? Must it be a human reading the data, or does computer analysis suffice?

2. How can, if it all, digital information be seized?

3. What makes obtaining digital information reasonable? Is it defined by the way in which the information is obtained, some characteristic of the information itself, some combination of the two, or something else entirely?

Each of these questions implicates a challenge for the interpretation of the privacy protection under the Fourth amendment. The following paragraphs will describe the way in which

the current legal landscape looks in the terms of these three questions.

Searches Most of the jurisprudence in privacy is actually the determination of whether or not a search occured. For a search to have occured, the information obtained must have been gotten unreasonably. The common test used here is if the person had a "reasonable expectation of privacy" for that information. It appears as though this expectation exists in one of three forms:

- The information was created, stored, and used only inside the home, thereby invoking the traditional protections of the private sphere of the home
- The information cannot be obtained by any common means and therefore cannot be expected to be ascertained, as in Kyllo
- Society recognizes the information as protected, as this expectation is recognized by the Court

Seizures The most common legal definition is that a seizure occurs when a person's "possessory interest" is affected by the action. Because the copying of digital data can occur without the inconvenience (or even the knowledge) of the owner of the information, digital information is almost never found to be seized.

Reasonableness When applied to the scenario of obtaining third party information, the only real inquiry into reasonableness is whether or not the entity in question went through the proper legal channels, as required by statute, to obtain the protected information. If not covered by statute, there is no reasonableness to consider because there is no search.

4.6 Outcomes

The situation created by the legal and technical circumstances described in this chapter is a system in which the government now pays the private sector to do exactly what they

themselves were prohibited from doing - amassing large databases of personal information to be stored until needed by the government. To summarize what occured:

- The computer and networking technology made it easier to to collect and share information
- This also allowed data mining technology, which then allowed new knowledge to be obtained from that information, which is useful both in private and public endeavors
- The mass adoption of such technologies, fueled both by legislation and society, led to massive digital dossiers being available of every person
- The legislature, concerned about unchecked power, passed laws trying to restrict government ability to collect massive amounts of information, but not regulating the flow and use of information otherwise
- To fill the desire for government to continue to take advantage of these technologies, data aggregation companies arose, now acting in the exact same role to hold mined information for the government to search at their will

These companies, like Choicepoint, Acxiom, LexisNexis, Docusearch, and many others are filling the role of collecting information and selling it almost completely unregulated. No longer is information protected simply by being infeasible to collect - it is all being collected and stored for later use. These actions are not illegal in and of themselves, but the current system poses several dangers to the individual.

Lack of awareness of what information about themselves is being collected Most typical users assume that information they choose to give to an organization will be used solely by that organization. For example, many people choose to become members of grocery store frequent shopper programs, which use savings cards. When the individual purchases items at the store, their card is scanned by the employee checking them out, and some discounts are applied to their items. However, in exchange for some savings, the individual is allowing the grocery store to associate that person with the items that he buys. People already give up

some privacy to this information each time they shop, because the other people in the store, particularly the checkout person, can see exactly what is in their cart. However, what the individual does not realize is that the company can compile that information over time, which can reveal very personal information (the inference problem). In addition, that grocery store may choose to sell the food profiles to other companies, who then combine it with information. Hypothetically, that information could be used to infer other pieces of information like religion, any changes in his life, or many other more personal facts or assumptions which are much larger than simply what food he ate.

In addition, the information can often be freely traded and aggregated many times, so the individual will probably never know who has it, how they use it, and if it is being sold to the government in some way.

Lack of awareness when or why information about them is being gathered or analyzed Historically, a person who was under investigation by the police tended to know about it. The police would have to talk to the suspect himself or people that knew him, who could likely tell the individual. Transparency in this regard helps ensure that the police do not abuse their investigatory power. However, with data mining, the subject never knows and can never tell if he is being unfairly targeted.

Inability to correct erroneous information Because the individual does not know everyone who possesses his information, it is impossible to find errors and correct them. Even if he knew the errors existed, it can be impossible to trace the trail of the personal information between the myriad of data miners who buy and sell personal information millions of times a day. That erroneous information about the individual exists in digital form on the internet, and it can no longer be recalled nor contained.

This is perhaps the most widely recognized shortcoming in today's society due to the prevelence of identity theft. When identity theft takes place, the person whose identity was used often experiences problems with erroneous information on his credit report, due to the fraudulent activity. However, getting this information corrected is a lengthy, expensive process

that is still not always effective at clearing the innocent party. It has been reported to take 150 to 500 hours of time and up to $3000 to fix an identity theft problem[32].

Inability to remain anonymous in daily activities The idea of privacy in public may sound counter-intuitive, but it is an assumption of daily life today. Everyday people walk out of their homes to go about their daily business. Most, if not all, realize that the casual passerby may notice what time they leave and what they are wearing. That information is external and public because it occurs in a public place. However, individuals expect that they will not be watched by the same person every day, giving the habitual observer insights into the habits and personal life of the individual that the casual observer simply cannot determine during isolated instances of notice. The idea of being repeatedly watched by someone is intrusive and perhaps uncomfortable, but the hypothetical can go even further. Now suppose that the individual lives in a city where there are cameras on every street corner that attempt to capture crime. Now the details of the individual's comings and goings around the entire town are recorded and aggregated. Technical improvements in facial recognition are starting to make it possible for the people collecting the footage from those cameras to identify individuals in each camera shot and then follow the individual's movements throughout the town.

Inability to choose which behaviors are worth the privacy risk The previous examples illustrate that even traditionally public activities are being used to collect massive amounts of information about private individuals in a way that was not possible before. Many individuals choose to put much of their private information out for public use, such as people who use new personal networking technologies like Facebook or Myspace to share details about themselves and their lives with many others. However, many people also choose not to be so open about their information. They may prefer to remain anonymous in their shopping and other personal habits. Unfortunately, privacy is not really an option anymore. It is increasingly difficult, probably impossible, to participate in today's digital society without relinquishing control of some personal information to a third party, and thereby making it available to the government.

CHAPTER 5. Discussion of Proposed Solutions

This chapter will examine various proposed solutions for the privacy problems described in the last chapters. Then, it will attempt to normalize all the different options and compare them using a 2D Stakeholder/Method Perspective Model. Finally, it will discuss the implications of that analysis.

5.1 Solution Analysis

5.1.1 The Model

A 2D Stakeholder/Method Perspective Model will be introduced to evaluate each of the solutions. The model I propose is based in large part from the work of Timothy Terrell and Anne Jacobs. In their analysis of two Supreme Court decisions with privacy impacts, they formulated a four square framework for evaluating the philosophies underlying judicial interpretations of the case involving privacy[39]. It was proposed as a means to better understand the way in which the competing interests were evaluated and weighed, which would also help in preparing arguments. It proposed two ways of categorizing the perspective of the opinion:

1. Moral philosophy, relating to the concept of fairness

2. Political philosophy, relating to the concept of justice

These categorizations led to four different combinations of these two perspectives describing the emphasis on which an argument is based, as seen in Figure 5.1[39]:

I propose to use a similar dual-dichotomy approach as a way of comparing the fundamental emphases of the solutions being evaluated. This approach will evaluate the solution in two ways:

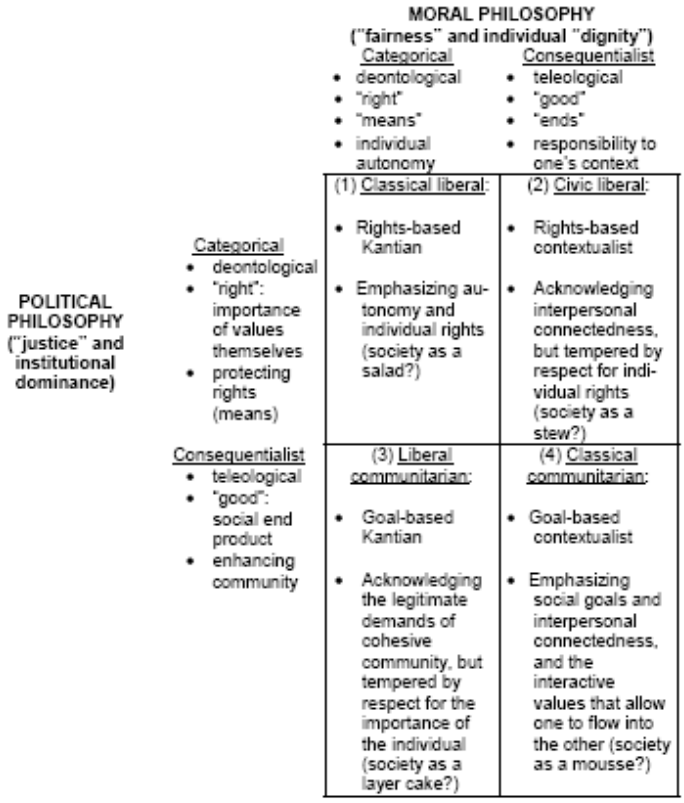

Figure 5.1 Reasoning Analysis

1. It will determine the stakeholder perspective, individual or societal, taken as the most important

2. Then it will evaluate the method perspective, ends or means, that is emphasized

To perform the comparative analysis, each of these evaluations will be placed on an axis and used to graph the points where various solutions would fall on these dichotomies. The model will be used to score the relative weight of these dichotomies on a scale of -5 to 5. Table 5.1 shows the guidelines used in assigning the scores.

At the end of each proposed solution description, a score for each criteria will be assigned and the rationale for that score will be explained.

Table 5.1 Solution Scoring

Score	Stakeholder Perspective	Methodology Analysis
-5	Inquiry hinges solely on the rights of the individual	Inquiry hinges solely on way data is obtained
-3	Strong favoring of rights of individual	Strong favoring of determining way data is obtained
-1	Inquiry favors slightly the individual over societal interest	Inquiry favors way data is obtained slightly
0	Equally weighs individual and societal interests	Equally weighs the way of gathering data with the end informational goal
1	Slightly emphasizes social perspective	Slightly emphasizes end goal over the method of collection
3	Strong favoring of societal perspective	Strong favoring of end goal over method of collection
5	Inquiry hinges solely on the societal interest	The information obtained, not the methodology, is the sole factor

5.1.2 Judicial Solutions

5.1.2.1 Restricted Third Party Doctrine (R3PD)

This approach would have the judiciary restrict third party doctrine to only apply to information deliberately conveyed in order that its content be used[11, 12]. Beyond that point, it would involve evaluating privacy claims on a case-by-case basis because it would be impossible to differentiate among all the possible types of information and distinctions in the way they are shared. Henderson cited the New Mexico appellate court, which rejected the third party doctrine as it currently stands, in saying

> In all cases that invoke [our Fourth Amendment analog], the ultimate question is reasonableness. We avoid bright-line, per se rules in determining reasonableness; instead, we consider the facts of each case.

The main argument in this view is that the 'reasonable expectation of privacy' of a subject is not automatically diminished by the sharing of this information with a third party, as the Court had declared in Smith v Maryland. Some state courts have found that society has come to expect a certain level of privacy in third party information, which may then invoke the a

reasonable expectation of privacy.

He proposed a set of factors used to determine reasonableness[11]:

1. The purpose of the disclosure
2. The personal nature of the information
3. The amount of information
4. The expectations of the disclosing party
5. The understanding of the third party
6. Postivite law guarantees of privacy
7. Government need
8. Personal recollections
9. Changing social norms and technologies

In addition he emphasized that some factors should not be used in anlysis and listed the following factors as irrelevant:

1. The form of the information
2. The 'good citizen' motivation of a third party
3. The government's method of acquisition
4. Expectation created by police conduct

Score The items enumerated in this list explicitly include government need and characteristics of the person involved. However, many more of the criterion focus on the impact on the individual. *Stakeholder Perspective = -2* The criteria listed specifically discount characteristics like the form of the information and the method of acquisition, while focusing on expectations of privacy between the parties. This shows a specific emphasis on the ends, rather than the means of collection. *Method Perspective = 3*

5.1.2.2 Informational Access Interpretation (IAI)

Manish Kumar suggests the Fourth amendment should be interpreted in respect to the inforation access[19]. He argues that it is not disclosure to public that makes information unprotected, but the ability of the public to meaningfully access and interpret that information that matters. He makes this argument primarily in the context of email, the contents of which are often publicly available to the ISP but are not needed as part of the business record in the way that email header information is needed. Therefore, because the ISP itself does not access the content, much like a post office could but does not open mail, the information should be protected.

Score The key analytical factor is the ability of the public to meaningfully access and interpret information, instead of an inquiry into the actions or invasions of the individual's realm. *Stakeholder Perspective = 3* The approach specifically asks the question: 'Does the government have to employ special *means* not available to the public to access the allegedly private information[19]?' If that answer is yes, it is a search. *Method Perspective = -5*

5.1.2.3 Seizure of Intangible Property (SIP)

Paul Ohm argues that the concept of what constitutes a seizure should be reverted to the 'dominion and control' definition rather than the current standard, which is characterized as 'meaningful interference with an individual's possessory interests in ... property[26].' He argues that this area of the law has not been pursued since the 1920s because in typical cases, before something can be seized a search must occur. Therefore, the definition of gathering information using new technology has hinged upon whether or not obtaining that information incurred a search, not whether gathering it is a seizure. However, advances in technology are making it possible to gather digital information without entering on any property and even without the user's notice. He argues that copying digital data dimishes the individual's dominion and control of the information, even though it does not interfere with their ability to possess the original copy of the information.

Score This perspective solely takes into account the perspective of the individual, relying only on his 'dominion and control' of his information. *Stakeholder Perspective = -5* The method of obtaining information is not the issue; instead the concern is the effect that obtaining that information has on the individual's rights. Therefore, the emphasis is strongly on the ends, rather than the means, of the collection. *Method Perspective = 4*

5.1.2.4 Human Access (HA)

Orin Kerr argues that the Fourth amendment regulates the flow of information between the individual and the government by limiting the access to information for human observation[17]. He stipulates that this means:

- Once information has been exposed to human observation, it is a search
- Copying data should not be a seizure, but copied data should be treated the same as the original in obtaining 'reasonableness' to search it. This means that changing the location and the format of the information should not inherently change its level of protection.
- This regulation does not limit the ability to understand information that it can access. Therefore, encryption does not incur a higher standard of protection. If this information can be accessed, then they have to right to try to break the encryption[16].

Score He argues that individual privacy is simply the byproduct of a larger societal goal: regulating government access to information. *Stakeholder Perspective = 2* The emphasis in Kerr's analysis is exactly what must occur to constitute a search or seizure. However, the emphasis on the impact of human observation implies that the end goal of maintaining privacy from agents of government is more important than the way in which that knowledge is reached. To explain this dissonance, I argue that Kerr implicates this end goal by saying that the purpose of the Fourth amendment is to regulate flow, and to do this he wants to regulate the way in which that access can be achieved. *Method Perspective = -1*

5.1.2.5 Plain Concealment (PC)

Jim Harper argues that the entirety of electronic records today are so revealing that they should be protected as the 'modern iteration of papers and effects[10].' In his interpretation, this would mean that concealing information from the *general* public is effectively concealing information from government. In doing so, a person would not have to exhibit any expectation for its protection, because the Constitution never requires an expectation of a right for it to be protected. Therefore, privacy would be a question of fact, rather than subjective analysis of societal expectation. This paradigm would be called 'plain conealment,' and very similar to 'plain view' exceptions in analysis.

Score Harper largely discounted the so-called 'chilling effects' of government surveillance on societal feelings, instead relying on individual rights and their efforts or happening to plainly conceal their information. *Stakeholder Perspective = -5* In making the attainment of information a fact-based analysis, Harper is inquiring into how exactly that information is obtained. If the individual had concealed the information, then attaining it must have required some kind of bypass. *Method Perspective = -5*

5.1.2.6 Long View (LV)

Lerner and Mulligan described an idea of taking the 'long view', or looking past the caveats of the current situation and asking two fundamental questions to determine the long term effects of the decision[14]:

1. Does the record which is held by the third party reveal information about activities that are taking place inside the home and would not be able to be otherwise obtained without physcial trespass into the home?

2. Is the consumer lacking any real choice about whether to create such records or an opportunity to choose to maintain them in the home?

If the answer to both these questions is yes, then obtaining this information should be protected by the Fourth amendment.

Score The two questions proposed in the framework itself makes scoring this example incredibly simple. The first question frames the information obtained as a characteristic of the individual, not societal expectation or influence. *Stakeholder Perspective = -5* The second question does not rely directly on how the information was obtained, but in what options the subject had to stop its release to the public. In taking this into consideration, the analysis is considering the mandatory loss of privacy of the individual before the way in which the information is collected. *Method Perspective = 3* It may also be argued that the overarching emphasis of this model is societal-ends focused, rather than individual-ends. Although I agree that the societal-ends seems to be a relevant concern in the explanation for the solution, the method for applying the questions would be emphatically more focused on the individual.

5.1.2.7 Proportionality Principle (PP)

Slobogin proposed another way to interpret the societal expectation of privacy from intrusion in his articulation of the proportionality principle. He thought that the intrusiveness of knowing information is a characteristic of the information, not how it is obtained. Therefore, the level of justification for search and seizure should be proportional to the intrusiveness of the information[35].

Score This perspective attempts a direct balance between the perspective of the individual and the need of law enforcement. In advocating proportionality, he is making a direct attempt to balance the two stakeholders. *Stakeholder Perspective = 0* As part of proportionality, the method of gathering information is important, but as that method is deemed appropriate by the ends. Again, the attempt was to balance the two, but with a slight emphasis on the ends. *Method Perspective = 1*

5.1.3 Legislative Solutions

5.1.3.1 Privacy Gap (PG)

The Center for Democracy and Technology proposes adapting the Privacy Act of 1974 to include government use of commercial databases[29]. Essentially, it argues that the government should not be able to rely on information from commercial databases unless they are subject to data quality and reliability standards.

Score Although the concept of individual privacy is hinted at throughout the analysis, the view here is from the impact of increasing government power in terms of its information collection abilities and the way that will impact society. *Stakeholder Perspective = 3* This article advocates exploring very specific ways to regulate the gathering and use of information. It recommends passing legislation that will govern exactly how government agents can obtain information about individuals. *Method Perspective = -3*

5.1.3.2 Regulate Personal Data (RPD)

In his article about data brokers and law enforcement, Chris Hoofnagle argues that historically, the legislature found it more important to protect against government intrusion rather than commercial use of personal data[13]. However, unchecked, the commercial entities have been using that distinction to collect the data and sell it to law enforcement. Hoofnagle recommends four changes to privacy legislation:

1. Minimize data collection
2. Remove distinctions between government and commercial collectors
3. Address emerging privacy issues with the exposure of personal information in public records
4. Amend the Privacy Act to apply to the sale of information to the government in the situations where the government is not allowed to collect it

Score Hoofnagle focuses on the role of privacy law to balance the power between individuals and the government. He also mentions that public records, which will lead to numerous individual privacy problems as they become digitized, serve as a way for society to observe the actions of officials - thereby increasing the ability of society to keep government abuse of power in check. His perspective recognizes the societal value in the issues, but strongly advocates increased protection of the individual. *Stakeholder Perspective = -2* He advocates specific limitations on the ways information can be collected and made public by the government (via public records like court documents). These foci rely on regulating the means of information collection. *Method Perspective = -2*

5.1.3.3 Degree of Accessibility (DA)

In Access and Aggregation, Solove is exploring the specific risks that governmental public records pose to privacy[36]. Although he frames his opening hypothetical situation in reference to individual information, he tends to refer to reforming public conceptions of privacy to mean increased control and limited access, not absolute secrecy. Solove argues that information does not have to be entirely private or not at all, but instead involves an expectation of a certain degree of accessibility. Specifically, he examined the way that the release and digitization of public records introduces a concern about the structure of information flow in society. He recommended nationalizing public records law and restricting some types of personal information from being published to the general public.

Score Solove's emphasis is slightly more about changing societal views of protection of personal information than the personal protection itself. By instituting a regime of degrees of accessibility, his solution would reduce the social reliance and acceptance of freely available personal information via public records. *Stakeholder Perspective = 1* This article advocates exploring very specific ways to regulate the gathering and use of information. It recommends passing legislation that will govern exactly how government agents can share and obtain information about individuals. *Method Perspective = -3*

5.1.3.4 Opt-In (OI)

Some current privacy legislation requires that companies allow individuals, if there is an option, to opt-out of the sharing of their information with nonaffiliated third parties. This process can be difficult, confusing, and many people are unaware that it even exists. Some proponents argue that customers should be able to opt-in to data sharing for additional benefits rather than be made to opt-out.

Score Advocates of this view primarily focus on the fact that individuals tend not to understand or choose to participate in opt-out systems. Instituting opt-in would make an individual's default status as protected, but the system of sharing information in society would not necessarily have to be impacted at all. *Stakeholder Perspective = -3* The opt-in system is a more ends-focused approach because it is enacted to make protecting personal information easier for individuals by reducing the need for action taken by the individual. It does not change the way that protected and unprotected information can be collected or handled, just simply reverses which lists would be kept by private companies. *Method Perspective = 3*

5.1.4 Societal Solutions

5.1.4.1 Pay for Privacy (PfP)

This scheme would create a business model where consumers interested in protecting privacy in commercial transactions could pay to not have their information revealed.

Score The pay for privacy scheme would allow individual users to choose the level of privacy they would prefer and weigh that value amongst others like cost or functionality. Although this is very similar to OI, it has greater societal emphasis than OI because this model relies on societal views to influence the emergence of this system. *Stakeholder Perspective = -1* This approach implies that companies who would have the power to collect would refrain in order to respect a user's privacy. The focus is not on regulating types of access, but on affording levels of privacy. Therefore, the primary focus of this model is the end goal, not the

method. *Method Perspective = 3*

5.1.4.2 Competitive Advantage (CA)

This scheme suggests that as privacy becomes a more important concern of consumers, companies will begin to attempt to distinguish themselves by placing a higher value on privacy of personal information. An interesting caveat of changes like the type that could evolve from privacy as a marketable value is that companies that begin to self-enact additional privacy restrictions could impact other organizations to do the same. As the EU Directive has caused companies from outside the EU to conform to its standards in order to do business with those companies under the directive, fostering a competitive privacy market within the US could do the same.

Score The approach takes a very total view of society and the market and views additional personal privacy as merely a commodity. *Stakeholder Perspective = 4* Like the pay for privacy model, the primary focus would be on the end goal of providing various levels of privacy, not regulating specific means of collection. *Method Perspective = 3*

5.1.5 Technical Solutions

The previous chapter identified three basic categories of technical solutions have been proposed to help protect privacy: identity hiding (IDH), information hiding (INH), and information minimization (IM). Every technology introduced provides the means to protect information, so each item in this category will have a Method Perspective score of -5.

Identity Hiding Score Hiding identity involves separating each individual from information about them, whether it is their browsing history or anonymity in a political poll. *Stakeholder Perspective = -5*

Information Hiding Score Hiding information not only protects the individual interest in keeping their information private, but also a societal interest in protecting a reasonable

balance of power between government need to police and individual autonomy. Although most of these technologies will have an impact primarily at the individual level, their existence has some social implications. *Stakeholder Perspective = -4*

Data Minimization Score Much like information hiding, these techniques attempt to protect the individual from oppression or interference from government. In doing so, it will provide some potential for societal benefit. *Stakeholder Perspective = -4*

5.2 Model Outcome Analysis

Determining which, if any, of these solutions should be enacted is a difficult decision, and not one that everyone is able to make. For example, individuals have little control over the drafting and passing of legislation, but they can contact their representatives and vote for people who are dedicated to protecting privacy. The only individuals who can enact new judicial approaches are judges; however, changes in societal values toward privacy and the use of information could increase the likelihood of a change in judicial review of the topic.

This thesis does not aim to recommend a particular solution or course of action as the 'best' to implement. To do so would require determining the appropriate balance of emphasis on different values, which is subject to personal opinion and interpretation. Instead, it hopes to provide a comparative analysis of various solutions to illuminate the types of changes that could be implemented and the effects it may have. In doing so, this may also prove to provide helpful insight into the way that new developments with privacy implications should be examined prior to their enactment.

This resulted in the following comparative scoring of the solutions, shown both in Table 5.2 and Figure 5.2

From the table and graph, it is possible to make several observations about the grouping of the data points:

1. Only one item plotted on the graph was anywhere near the data point 'Now', which is indicating the current interpretation of privacy when this model is applied. The nearby

Table 5.2 2D Stakeholder/Method Perspective Scores Table

Solution	Stakeholder Perspective	Method Perspective
Now	4	2
Judicial Solutions		
R3PD	-2	3
IAI	3	-5
SIP	-5	4
HA	2	-2
PC	-5	-5
LV	-5	3
PP	0	1
Legislative Solutions		
PG	3	-3
RPD	-2	-2
DA	1	-3
OI	-3	3
Societal Solutions		
PfP	-1	3
CA	4	3
Technical Solutions		
IDH	-5	-5
INH	-5	-4
IM	-5	-4

item was CA, referring to the societal measure of allowing competitive advantage. Every other solution proposed was in a different quadrent.

2. All the technical solutions are very focused at the individual/means level.

3. All but one of the legislative solutions were at roughly the same means level as well. Although the legislative aims varied in individual and societal emphasis, the regulation took place at a very detailed level invoking the means of collecting and using information.

4. Judicial interpretations varied wildly in data point placement.

5. Only one point was on an axis, and very few were close. Most solutions strongly favor one perspective over the other.

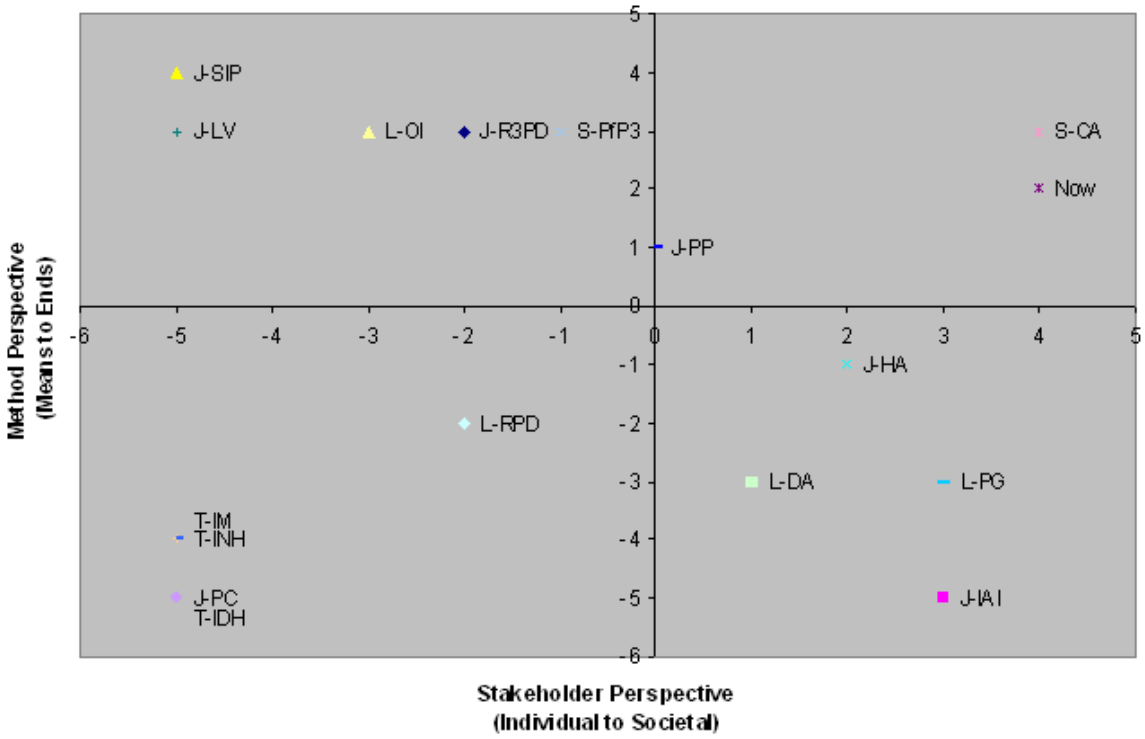

Figure 5.2 2D Stakeholder/Method Perspective Data Plot

5.3 Implications of the Findings

Analyzing the proposed solutions, which attempt to solve the same issues in different ways, revealed interesting trends as well as potential gaps. This section will attempt to interpret the deeper meanings of the plot analysis provided in the previous section. The key findings are:

1. Some combination of legislative and technical solutions will be the most likely candidates for regulating the means by which information is collected and analyzed. However, these solutions will be ever evolving because technology and society will continue to change. However, the nature of law and technology is that they are often effective at incremental change. Laws and technologies are capable of, but rarely do, producing radical change. Instead, they will be the slow way that checks and balances between the individual and government are enacted.

2. The potential impact of a judicial paradigm shift in interpreting the Fourth amendment could be huge and very unpredictable. The many methodologies evaluated in this paper would all revamp some aspect of Fourth amendment interpretation, primarily in either the interpretation of the third party doctrine, the reasonable expectation of privacy definition of a search, and the possessory interest definition of a seizure. Any of these changes will have sweeping implications in the direction that society is taking. Any of these changes will likely lead to a massive restructuring of government information attainment strategies. In addition, altering the third party doctrine could destroy or severely reduce the exploding information aggregation and sales market, particularly for those companies that target law enforcement and government agencies as their primary customers.

3. Solutions that balance these competing interests, individual versus society and means versus end, make up only a small percentage of the proposed solutions. Many of these solutions very strongly favored one quadrant instead of having balance on either dichotomy. Does this mean that we must choose one set of values over the other? To some degree, I believe we do. The only solution that was on an axis was Slobogin's Proportionality Principle doctrine. The problem with this doctrine would be its implementation. Because it does not settle the struggle between the two opposing forces (government and individuals) this battle will be rehashed repeatedly as the Court attempts to interpret the complex technical, legal, and societal factors. This solution is what I would call a high-risk, high-reward. It has the potential to give the Court the flexibility to interpret each situation in a complete and meaningful way to determine reasonableness under the Fourth amendment. However, because it is so ambiguous, this solution is also the most likely to lead to a future of unpredictable Court opinions and then unclear societal expectations about privacy.

5.4 Conclusion

Performing a comparative analysis of all the proposed solutions across different disciplines helped shed light solution landscape and the ways in which privacy could be impacted in the future. The 2D Stakeholder/Method Perspective data plot clearly depicted the solution landscape and provided useful means to look for new information. With the data points, I was able to spot trends or lack thereof, look for weaknesses, and hypothesize about the future. Hopefully this analysis will serve as a useful framework to evaluate potential solutions and provide a way to ponder potential new solutions that have not yet been considered.

BIBLIOGRAPHY

[1] Arendt, Hannah. *Between Past and Future*. New York: Viking Press, 1961. Revised edition, 1968.

[2] Bellinger, Gene, Durval Castro, and Anthony Mills. Data, Information, Knowledge, and Wisdom. *Systems Thinking*. 2004. 21 March 2009. `http://www.systems-thinking.org/dikw/dikw.htm`

[3] Brin, David. *The Transparent Society*. Reading, MA: Addison-Wesley, 1998.

[4] Chopra, S., and White, L. Privacy and artificial agents, or, is google reading my email? In Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07) , (pp. 1245-1250). AAAI Press. 2007.

[5] Conti, Greg. *Googling Security*. Boston: Addison-Wesley, 2009.

[6] Credit Report Errors are Very Common. FreeCreditReportGuide.org. 16 March 2009. `http://www.freecreditreportguide.org/free-report/errors.php`

[7] DIKW. 18 March 2009. *Wikipedia*. 21 March 2009. `http://en.wikipedia.org/wiki/DIKW`

[8] Farkas, Csilla and Sushil Jajodia. The Inference Problem: A Survey. *ACM SIGKDD Explorations Newsletter*,Volume 4 , Issue 2. December 2002.

[9] Garfinkel, Simson. *Database Nation: The Death of Privacy in the 21st Century*. Sebastopol, CA: O'Reilly and Associates, Inc.: 2000.

[10] Harper, Jim. Reforming Fourth Amendment Privacy Doctrin. *American University Law Review*, Volume 57: 1381, 2008. Available at: http://www.wcl.american.edu/journal/lawrev/57/harper.pdf?rd=1

[11] Henderson, Stephen E.,Beyond the (Current) Fourth Amendment: Protecting Third-Party Information, Third Parties, and the Rest of Us Too. *Pepperdine Law Review*, Vol. 34, 2007; Widener Law School Legal Studies Research Paper No., 08-11. Available at SSRN: http://ssrn.com/abstract=922343

[12] Henderson, Stephen E. Nothing New Under the Sun? A Technologically Rationale Doctrine of Fourth Amendment Search *Mercer Law Review* Vol. 56, No. 507, 2005. Available at SSRN: http://ssrn.com/abstract=722125

[13] Hoofnagle, Chris Jay. Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect, Process, and Package Your Data for Law Enforcement. 29 NCJ Int'l L and Com. Reg. 595. Summer 2004.

[] Information Awareness Office. *Wikipedia*. 1 April 2009. http://en.wikipedia.org/wiki/Information_Awareness_Office

[14] Lerner, Jack I. and Deirdre K. Mulligan. Taking the 'Long View' on the Fourth Amendment: Stored Records and the Sanctity of the Home. *Stanford Technical Law Review*, 3, 2008. Available at: http://stlr.stanford.edu/pdf/lerner-mulligan-long-view.pdf

[15] Lyman, Peter and Hal R. Varian, How Much Information, 2003. 18 March 2009. http://www.sims.berkeley.edu/how-much-info-2003.

[16] Kerr, Orin S. The Fourth Amendment in Cyberspace: Can Encryption Create a Reasonable Expectation of Privacy?. *Connecticut Law Review*, Vol. 33, p. 503, 2001; GWU Law School Public Law Research Paper No. 219; GWU Legal Studies Research Paper No. 219. Available at SSRN: http://ssrn.com/abstract=927973

[17] Kerr, Orin S. Searches and Seizures in a Digital World. *Harvard Law Review*, Vol. 119, 2006; GWU Law School Public Law Research Paper No. 135. Available at SSRN: `http://ssrn.com/abstract=697541`

[18] Kerr, Orin S. A User's Guide tot the Stored Communications Act - and a Legislator's Guide to Amending It. *George Washington Law Review*, 2004. Available at SSRN: `http://ssrn.com/abstract=421860` or DOI: 10.2139/ssrn.421860

[19] Kumar, Manish. Constitutionalizing Email Privacy by Information Access. *Minnesota J.L. Science and Technology* 9(1), 257-286. 2008.

[20] McArthur, Robert L. Reasonable Expectations of Privacy. *Ethics and Informtion Technology* Vol. 3: 123-128, 2001.

[21] McCoy, M. K. In government we trust: exchanging information for public health services. Paper presented at the annual meeting of the American Association for Public Opinion Research, Sheraton Music City, Nashville, TN. 16 August 2003. 6 February 2009 from `http://www.allacademic.com/meta/p116339_index.html`

[22] Mulligan, Deirdre K., Ari Schwartz, and Indrani Mondal. Risks of Online Storage. *Communications of the ACM*, Vol. 49, No. 8. August 2005.

[23] Nissenbaum, Helen. Protecting Privacy in an Information Age: The Problem of Privacy in Public. *Law and Philosophy*, 17: 559-596, 1998.

[24] OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. Organization for Economic Co-operation and Development. 23 September 1980.

[25] O'Harrow, Robert, Jr. *No Place to Hide*. New York: Free Press, 2005.

[26] Ohm, Paul. The Olmsteadian Seizure Clause: The Fourth Amendment and the Seizure of Intangible Property. *Stanford Technical Law Review* Rev. 2, 2008.

[27] Palace, Bill. Data Mining: What is Data Mining? *Data Mining*. Spring 1996. Anderson Graduate School of Management at UCLA. 19 March 2009. http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm

[28] Privacy Laws by State. *Electronic Privacy Information Center*. 1997. 22 March 2009. http://epic.org/privacy/consumer/states.html

[29] Privacy's Gap: The Largely Non-Existent Legal Framework for Government Mining of Commercial Data. *Center for Democracy and Technology*. 28 May 2003. Available at: http://www.cdt.org/security/usapatriot/030528cdt.pdf

[30] Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment. Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals, National Research Council. 2008.

[31] Reidenberg, Joel R. *E-Commerce and Trans-Atlantic Privacy*, 38 Hous. L. Rev. 717, 730 (2001).

[32] Schmidt, Steffen and Michael McCoy. *The Silent Crime: What You Need to Know about Identity Theft*. USA: Twin Lakes Press, 2008.

[33] Schneier, Bruce. Why Data Mining Won't Stop Terror. *Wired*. 9 March 2006. 15 March 2009. http://www.wired.com/politics/security/commentary/securitymatters/2006/03/70357

[34] Sheer, Robert. Nowhere to Hide. *Yahoo Internet Life* (Special Report on Privacy), 6(10): 101, October 2000.

[35] Slobogin, Christopher. Government Data Mining and the Fourth Amendment. *University of Chicago Law Review*, Forthcoming. Available at SSRN: http://ssrn.com/abstract=1001972

[36] Solove, Daniel. Access and Aggregation: Public Records, Privacy, and the Constitution. *Minnesota Law Review*, Vol. 86, No. 6, 2002. Available at SSRN: http://ssrn.com/abstract=283924 or DOI: 10.2139/ssrn.283924

[37] Solove, Daniel. A Brief History of Information Privacy Law. *Information Privacy Law*. 2d ed 2006.

[38] Solove, Daniel. *The Digital Person*. New York: New York University Press, 2004.

[39] Terrell, Timothy P. and Anne R. Jacobs. Privacy, Technology, and Terrorism: Bartnicki, Kyllo, and the Normative Struggle Behind Competing Claims to Solitude and Security. *Emory Law Journal*. Fall 2002.

[40] Thearling, Karl. An Introduction to Data Mining: Discovering Hidden Value in Your Data Warehouse. 15 March 2009. http://www.thearling.com/text/dmwhite/dmwhite.htm

[41] United Nations General Assembly. Universal Declaration of Human Rights. 10 December 1948.

[42] US Department of Health, Education, and Welfare. Records, Computers, and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems 29. 1973.

[43] Waldo, James, Herbert S. Lin, and Lynette I. Millett, eds. *Engaging Privacy and Information Technology in a Digital Age*. Washington, D.C.:The National Academies Press, 2007.

[44] Warren, SD and Brandeis, LD. The right of privacy. *Harvard Law Review* (December 1890), 193-220.

[45] Westin, Alan. *Privacy and freedom* (Fifth ed.). New York, U.S.A.: Atheneum, 1968.

[46] Who uses Tor? The Tor Project, Inc. 4 January 2009. 18 March 2009. http://www.torproject.org/torusers.html.en